

# Segmentation de corrélations par débruitage pour explorer l'intégration financière internationale

Pierre BORGNAT<sup>1</sup>, Cécile BASTIDON<sup>2</sup>, Pablo JENSEN<sup>1</sup>, Patrice ABRY<sup>1</sup>, Antoine PARENT<sup>3</sup>

<sup>1</sup>Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France

<sup>2</sup>LEAD, Université de Toulon, France

<sup>3</sup>LAET, Sciences Po Lyon (UMR CNRS 5593), & OFCE, France

{pierre.borgnat,pablo.jensen,patrice.abry}@ens-lyon.fr, bastidon@univ-tln.fr,  
antoine.parent@sciencespo-lyon.fr

**Résumé** – Nous étudions dans ce travail les changements de la structure de corrélation des séries de rendements des marchés d'actions dans le monde de 1960 à 2018, afin d'analyser de manière endogène quelles sont les périodes d'intégration financière internationale. Pour cela, nous considérons différents indicateurs décrivant les corrélations (évoluant dans le temps) et utilisons une méthode de débruitage de signaux multivariés constants par morceaux pour réaliser la segmentation.

**Abstract** – We are studying the changes in the correlation structure of multivariate series of log-return of world-wide equity markets, on the period 1960-2018. The goal is to analyse from endogenous data what are the different segments of international financial integration. For that, we consider different possible features describing the time evolving correlations and we leverage a denoising method for multivariate piecewise constant signals so as to obtain segmentations for analysis.

## 1 Introduction

L'objet de ce papier<sup>1</sup> est d'étudier l'intégration financière mondiale et pour cela d'identifier les événements qui conditionnent les changements de structure des corrélations entre rendements boursiers dans le monde depuis 1960. Nous donnons comme définition de l'intégration financière le fait que des bourses de plusieurs pays suivent globalement les mêmes évolutions et sont donc corrélées [3, 11, 12]. À cette fin, nous employons une approche issue de la théorie du signal, plus précisément nous nous appuyons sur la méthode de débruitage multivariée proposée dans [6]. Cette méthode vise à débruiter des signaux multivariés supposés être constants par morceaux avec des sauts en petit nombre aux mêmes endroits (éventuellement sans sauts pour certaines séries).

**Question économique étudiée.** Le problème macroéconomique est de localiser le changement dans la structure des covariances des rendements des marchés d'actions dans le monde depuis 1960 et ainsi d'identifier les plateaux qui correspondent aux périodes de stabilité de cette structure. Il est abordé par une approche de signal car nous cherchons une méthode de segmentation endogène. C'est-à-dire que l'on va partir des mesures des corrélations entre les rendements, dans une démarche de cliométrie, en s'in-

téressant à travailler sur des temps longs tout en partant des mesures au cours du temps avec des méthodes informées par les données. En histoire économique, il est plus courant de s'appuyer soit sur une datation d'événements de régulations économiques [11, 12] ou de tenter de déterminer, pays par pays, si il y a des convergences vers une intégration financière ce qui oblige à fixer des seuils un peu arbitraires de corrélation [3].

**Originalité de l'approche.** Les données dont on dispose (voir la section 2) sont les corrélations des rendements des bourses principales d'un ensemble de 32 pays, de 1960 à novembre 2018. Contrairement aux approches classiques de segmentation de covariance nous travaillons sur des indicateurs globaux décrivant l'évolution des distributions des corrélations paires à paires entre les rendements boursiers d'un ensemble de pays. En effets ces méthodes supposent en général la stationnarité par morceaux de la matrice de corrélation elle-même. Or ce n'est pas vraiment adapté ici puisqu'un pays ou un groupe de pays pourrait ne pas du tout suivre les mêmes évolutions (exemple typique : l'économie chinoise). Ainsi, nous allons proposer une méthode différente, qui ne s'appuie pas sur des hypothèses aussi fortes. L'autre difficulté tient à la nature des données : elles sont censurées à gauche car pour certains pays, nous ne disposons pas de données dès 1960, si leurs places financières de référence n'existaient pas (par exemple la Russie avant 1995) ou si elles ne sont pas ren-

1. **Remerciements** : travail réalisé dans le projet ACADEMICS de IDEXLYON de l'Université de Lyon dans le cadre du Programme Investissements d'Avenir (ANR-16-IDEX-0005), et à l'IXXI.

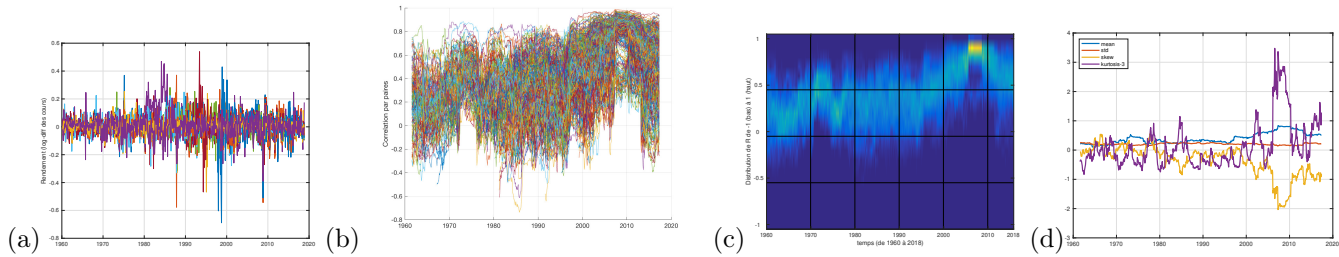


FIGURE 1 – Visualisation (a) des séries corrigées et log-différenciées (log-rendements), (b) des corrélations paires à paires, (c) de l’histogramme au cours du temps de ces corrélations et (d) des indicateurs moyenne  $\mu$ , écart type  $\sigma$ , asymétrie  $s$ , et aplatissement  $\gamma$ , à chaque  $t$ .

seignées auprès de la base de données FRED qui sert de référence. Pour toutes ces raisons, les méthodes usuelles ne sont pas adaptées ici : celles dans [8, 9]) qui segmentent la corrélation dépendante du temps, ou la matrice de précision [10], avec un cadre bayésien (e.g., [13]), ou avec des procédures d’optimisation spécifiques pour des pénalités sur les changements dans la structure de covariance [1] et qui peuvent servir pour des données financières [13], ou bien sur du télétrafic Internet [7].

## 2 Données de référence

**Sources des données et corrections.** Les séries temporelles dont nous partons sont les données mensuelles référencées à la FRED (Federal Reserve Economic Data)<sup>2</sup> sur les places boursières principales de 32 pays, couvrant tant des économies avancées, qu’en développement ou émergentes, de 1960 à nos jours (en 1960, seuls 15 pays sont représentés, et leur nombre augmente peu à peu jusqu’à arriver à 32 dans le panel à partir de 1999). Ensuite, tous les indices, exprimés en monnaie nationale, sont convertis en dollars à partir des taux bilatéraux entre les devises correspondant aux indices et le dollar US. Toutes les séries temporelles d’indice converties en dollars US sont ramenées alors en dollars constants base 2015, à partir de la série d’indice des prix aux Etats-Unis, de manière à les corriger de l’inflation.

Enfin ces séries temporelles corrigées du change et de l’inflation sont converties en séries de rendement par log-différenciation et ainsi rendues suffisamment stationnaires sur des courts temps (voir la en Fig 1 (a) pour un examen visuel) pour pouvoir estimer leurs corrélations  $\mathbf{R}(t)$  sur des fenêtres glissantes ;  $\mathbf{R}_{ij}(t)$  est alors le coefficient de corrélation de Pearson entre deux pays  $i$  et  $j$ , sur une fenêtre glissante centrée en  $t$  de  $w$  mois de long. Dans un premier temps,  $w = 36$  mois (soit 3 ans). Pour calculer une matrice de corrélation de 32 pays, il semblait nécessaire de ne pas prendre  $w$  trop petit même si cette valeur ne permet pas de dire que  $\mathbf{R}(t)$  est convenablement estimée en tant que matrice. Les méthodes utilisant des ma-

trices de corrélation correctement estimée, comme [8, 9] ne sont pas employables. Nous considérons ici les corrélations par paire individuellement, sans exploiter la structure de matrice de  $\mathbf{R}(t)$ .

**Etude qualitative des données.** Si l’on examine les séries des corrélations, en Fig 1 (b), on constate des fortes évolutions peu stationnaires au cours du temps. L’évolution qui guide notre intuition est celle de la figure 1 (c) montrant l’histogramme glissant des corrélations paires : les périodes d’intégration financière se traduisent par une distribution raisonnablement invariante dans le temps de ces corrélations. Par exemple, cela semblerait le cas autour de 2010, avec une période où les corrélations semblent toutes hautes et resserrées. Cependant, une période n’est pas forcément caractérisée par une convergence en moyenne comme à ce moment-là. Entre environ 1980 et 1994, les histogrammes semblent similaires d’un temps à l’autre car avec le même type de distribution large autour d’une moyenne donnée. La méthode employée va viser en quelque sorte à segmenter cette représentation de  $\mathbf{R}(t)$ , comme elle apparaît à travers son histogramme.

## 3 Méthode

**Indicateurs.** On testera trois possibilités dans la suite :

- (A) Employer directement l’ensemble des corrélations par paires,  $\mathbf{R}_{ij}(t)$ , soit 496 séries temporelles. Une difficulté est de correctement gérer les données absentes, en plus d’avoir un problème d’assez grande dimensions (496 séries de  $T = 670$  points en temps).
- (B) Employer les histogrammes normalisés de ces  $\mathbf{R}_{ij}(t)$  ; par exemple, on emploiera comme en fig. 1 (c) un histogramme avec des classes centrées de -1 à 1, par saut de 0.1.
- (C) Ne retenir que les premiers descripteurs des distributions des  $\mathbf{R}_{ij}(t)$  : leurs moyenne  $\mu$ , écart type  $\sigma$ , asymétrie  $s$ , et aplatissement normalisés  $\gamma$  (= kurtosis-3, donc 0 pour une gaussienne), à chaque  $t$ . L’avantage est d’avoir un nombre de séries stable pendant

2. <https://fred.stlouisfed.org>

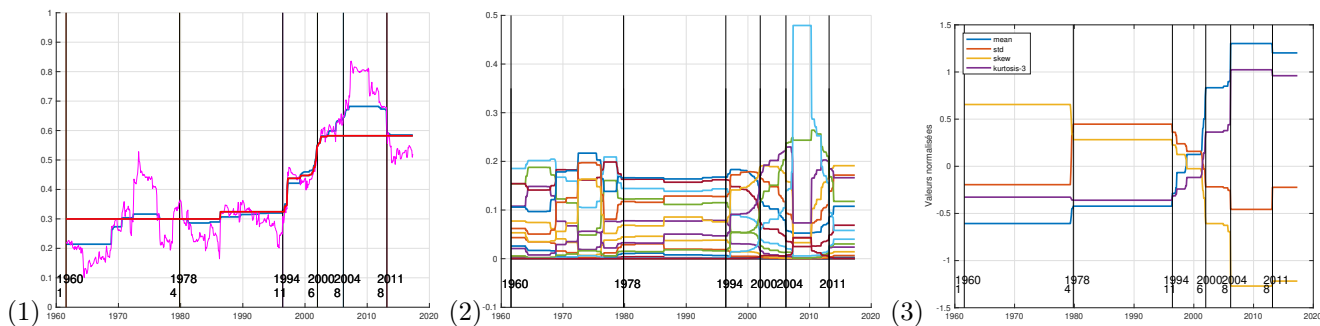


FIGURE 2 – Visualisation des segmentations obtenues en résolvant l’éq. 1 en partant : (A) de toutes les 96 séries  $\mathbf{R}_{i,j}(t)$  (où on a représenté 2 choix de  $c$  et  $\mu(t)$  en trait fin) ; (B) des histogrammes normalisés de ces corrélations (chaque courbe étant une des 21 classes) ou (C) des indicateurs  $\mathbf{X}(t) = [\mu(t), \sigma(t), s(t), \gamma(t)]$ . Les barres noires verticales visualisent les limites des périodes proposées par la segmentation des indicateurs (3) (attention, le décalage avec les axes n’est qu’apparent : les données sont centrées au milieu des fenêtres de 3 ans donc la barre en 1978/4 est alors quasiment affichée en 1980 puisque la fenêtre de 1980/01 commence en 1978/7). Voir le texte pour la discussion et l’interprétation.

toute la période d’étude, tout en ayant moins d’arbitraire que dans le cas (B) puisqu’on n’a pas à choisir l’échantillonnage des bins de l’histogramme.

**Débruitage de signaux multivariés constant par morceaux.** Notons  $\mathbf{X}(t)$ , pour les  $t = 1 \dots T$  les indicateurs retenus (par exemple  $\mathbf{X}(t) = [\mu(t), \sigma(t), s(t), \gamma(t)]$  dans le cas (B)). Le débruitage est fait en minimisant une fonctionnelle qui traduit un équilibre entre une attache aux données et un terme qui privilégie un signal constant par morceaux avec un nombre de sauts parcimonieux :

$$\arg \min_{\mathbf{U}} \|\mathbf{X} - \mathbf{U}\|^2 + c \sum_t \sqrt{\left( \sum_k |U_k(t+1) - U_k(t)| \right)^2}. \quad (1)$$

Le deuxième terme est une variation totale avec une norme  $l^1$  sur la dérivée qui, combinée avec la norme  $l^2$  sur l’attache aux données, va favoriser de ne retenir qu’un nombre réduit de sauts dans tous les signaux, aux mêmes instants. Le paramètre  $c$  règle ce compromis et contrôle indirectement le nombre de sauts (typiquement  $c$  est entre 10 et 100 dans les cas considérés). La minimisation de ce problème convexe non lisse est faite par un algorithme proximal primal-dual décrit dans [6], adapté du travail [5].

## 4 Résultats et interprétation

Pour les trois choix possibles d’indicateurs, on applique la méthode de débruitage aux données décrites en section 2. Notez qu’on peut déjà empiriquement caractériser les corrélations : dans les années 60, les corrélations sont légèrement croissantes ; dans les 70, avec l’augmentation du panel de pays dont des économies en développement, elles sont divergentes et plus étalées. Pour les décennies 80 et 90, elles sont croissantes et convergentes (toujours avec une panel qui augmente). Sur 2001-2008, elles sont stables et bi-polarisées (i.e., majoritairement élevées : autour de

0,7 à 0,9, mais avec un sous-groupe minoritaire de corrélations faibles : autour de 0,3 à 0,5). Enfin, post 2008 elles sont décroissantes, la bi-polarisation n’apparaît plus.

Pour les 3 choix possibles d’indicateurs, on ajuste le paramètre  $c$  de sorte à retenir un petit nombre de périodes stables, i.e., de morceaux constants en résolvant l’éq. (1), de manière à pouvoir interpréter la segmentation. Les segmentations les plus parlantes sont présentées en fig 2.

**Critique des segmentations par choix (A) et (B).** Pour la segmentation directe de la corrélation (A) : on obtient en fait le mode dominant lié principalement à la moyenne seule. Par exemple, en rouge (avec  $c$  plutôt grand), on distingue essentiellement deux périodes avec moins d’évolution, une avant 1994 et l’autre après 2002, mais on ne capte pas les effets de convergence (changement vers 1980) ou la modification de la bi-polarisation.

En segmentant sur l’histogramme, choix (B), on peine à obtenir des segmentations autres que fines, avec beaucoup de segments. En effet, ici la période entre 2004 et 2010 avec des corrélations élevées (donc des finances intégrées) sort bien mais si on augmentait  $c$ , c’est d’abord ce pic qui disparaîtrait (car  $c$  qui augmente réduit les valeurs estimées, du fait du biais intrinsèque de l’estimateur de débruitage). Obtenir des périodes stables plus interprétables gommerait les différentes périodes entre 2000 et 2010.

**Interprétation de la segmentation sur les descripteurs des histogrammes, (C).** Elle semble mieux fonctionner, car elle retrouve et quantifie les évolutions qu’on voit qualitativement. Ce sont celles qui sont reportées en fig 2 et nous les commentons plus avant ici en la mettant en regard avec des événements économiques liés aux ruptures de séquence de politique monétaire des principales banques centrales mondiales.

Cette segmentation obtenue par ces descripteurs de l’histogrammes à l’aide des 4 moments (normalisés et centrés) coïncide raisonnablement avec les tendances observées sur les histogrammes. La table 1 met en regard les segments

Période,	$[\mu, \sigma, s, \gamma]$	Événement économiques (Fed US et BCE)
60-78/4	$[-, -, +, -]$	Taux US croissants et fortement volatiles
78/4-94/11	$[-, +, +, -]$	Très fort relèvement des taux US, suivie de 15 ans de baisse
94/11-00/6	$[0, 0, -, 0]$	Période de stabilisation des taux US, niveau intermédiaire
00/6-04/8	$[+, 0, 0, +]$	Première période d'abaissement des taux US au « plancher de taux »
04/8-11/8	$[+, -, -, +]$	Brève sortie (2004-2007) puis retour des taux US du « plancher de taux »
Post 11/8	$[+, 0, -, +]$	Taux US au « plancher de taux ». Décisions de taux de la BCE inhabituels

TABLE 1 – Table synthétique sur la segmentation obtenue par débruitage des indicateurs (C) :  $[\mu, \sigma, s, \gamma]$ . Pour ces indicateurs, on indique +, 0 ou – selon que sur cette période il est plutôt grand, de valeur moyenne ou petit par rapport aux valeurs usuelles qu’il prend sur toute la période. Ces valeurs sont des descripteurs de la forme de la distribution.

obtenus avec des valeurs schématiques des 4 descripteurs et avec des événements économiques principaux décrivant la période. C’est ici particulièrement des décisions de la Fed (banque fédérale US) qui fixent ses taux directeurs et qui semblent dicter les segments à un même niveau d’intégration financière. Le comportement lors de la dernière période semble affecté par la décision de baisse des taux directeurs de la BCE depuis 2011 à des niveaux historiquement très bas. Notre clé de lecture de l’intégration financière et des événements majeurs sur les marchés d’actions mondiaux depuis 60 ans est que ce sont finalement des événements de politiques monétaires internationales, tels que les changements des taux directeurs de la Fed, qui conditionnent le comportement des places financières.

**Variations.** Nous avons exploré si la segmentation est robuste de diverses façons : i) en abaissant la taille de la fenêtre à 2 ans (24 points en temps, au lieu de 36) et ii) quand on retire des pays, par exemple les pays très périphériques au sens où leurs corrélations sont faibles avec les autres pays (ex. : la Chine ou Israël). La segmentation avec les descripteurs (C) reste quasiment inchangée.

**Conclusion.** Cette étude permet, en s’appuyant sur une expertise en théorie de signal et une démarche d’analyse exploratoire, de discriminer entre tous les candidats potentiels à l’explication des changements dans la structure des corrélations des rendements d’actifs, tels que : existence de crises financières, agrégat monétaire, agrégat de crédit, ou séquence de politique monétaire identifiée par les décisions de taux directeur des banques centrales. C’est ce dernier qui se révèle pertinent, puisque la méthode révèle que la politique monétaire américaine, sous l’effet de la modification du taux directeur de la Fed, coïncide avec tous les changements de structure des corrélations sur les marchés d’actions mondiaux, observés depuis 1960. Ce résultat met ainsi en évidence un canal à la Borio-Zhu [4] de transmission des chocs de taux directeur de la Fed sur la structure des marchés d’actions dans le temps.

Méthodologiquement, l’apport de la communication est de montrer qu’on peut explorer la segmentation d’une matrice de corrélation, même assez mal estimée et avec des données manquantes, en s’appuyant sur du débruitage de descripteurs de l’histogramme des corrélations. L’analyse du signal apparaît ici comme un outil cliométrique intéres-

sant pour détecter les déformations de structure des corrélations sur de longs temps. Un travail complémentaire reprend cette discussion avec des indicateurs réseaux pour caractériser la matrice de corrélation [2]

## Références

- [1] D. Angelosante and G. B. Giannakis. Sparse graphical modeling of piecewise-stationary time series. In *2011 IEEE ICASSP*, pages 1960–1963, May 2011.
- [2] C Bastidon, A Parent, P Jensen, P Abry, and P Borgnat. Graph-based era segmentation of international financial integration. *ArXiv :1905.11842*, May 2019.
- [3] G Bekaert, CR Harvey, CT Lundblad, and S Siegel. What segments equity markets? *The Review of Financial Studies*, 24(12) :3841–3890, 2011.
- [4] C Borio and H Zhu. Capital regulation, risk-taking and monetary policy : A missing link in the transmission mechanism? *J. of Financial Stability*, 8(4) :236 – 251, 2012.
- [5] L. Condat. A direct algorithm for 1D total variation denoising. *IEEE Sig. Proc. Lett.*, 20(11) :1054–1057, 2013.
- [6] J. Frecon, N. Pustelnik, P. Abry, and L. Condat. Fast and on-the-fly approximation of multivariate total variation minimization. *IEEE T. Sig. Proc.*, 64(9) :2355–2364, 2016.
- [7] A.J. Gibberd and J.D.B. Nelson. High dimensional changepoint detection with a dynamic graphical lasso. In *2014 IEEE ICASSP*, pages 2684–2688, May 2014.
- [8] A.J. Gibberd and J.D.B. Nelson. Estimating dynamic graphical models from multivariate time-series data : Recent methods and results. In *AALTD 2015. LNCS*, volume 9785. Springer, Cham, 2016.
- [9] D. Hallac, P. Nystrup, and S. Boyd. Greedy Gaussian Segmentation of Multivariate Time Series. *ArXiv :1610.07435*, October 2016.
- [10] D Hallac, S Vare, SP Boyd, and J Leskovec. Toeplitz inverse covariance-based clustering of multivariate time series data. *ArXiv*, abs/1706.03161, 2017.
- [11] M Obstfeld and AM Taylor. *Global capital markets : integration, crisis, and growth*. CUP, 2004.
- [12] K H O’Rourke and J G Williamson. *Globalization and history*, 1999.
- [13] X Xuan and K Murphy. Modeling changing dependency structure in multivariate time series. In *ICML*, pages 1055–1062, New York, NY, USA, 2007. ACM.